

PELABELAN PERAN SEMANTIS MENGGUNAKAN TATA BAHASA KASUS

SEMANTIC ROLE LABELING USING CASE GRAMMAR

Nur Indrawati¹ , M. Arif Bijaksana² , dan Ririn Dwi Agustin³

¹Jurusan Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom, Jl. Telekomunikasi Terusan Buah Batu Bandung, Jawa Barat, ³Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pasundan, Jl. Setiabudi No. 193 Bandung, Jawa Barat
e-mail: nurindrawati@gmail.com¹, email: arifbijaksana@gmail.com², e-mail: riryno@yahoo.co.id³

Naskah diterima tanggal 24 September 2013, direvisi tanggal 22 Oktober 2013, disetujui tanggal 8 November 2013

Abstract

The Application of semantic role labeling as one of the tasks from natural language processing include a variety of fields, such as machine translation, text summarization, and text mining. Semantic roles labeling is used in other text mining as the preprocessing of the concept-based mining system, which utilizes graphical representation and statistical-based data to find the concept of a sentence in the document text. Searching the concept on the preprocessing of text mining are one of the way to get valuable information from the sentence, taking into account the semantic of every word in a sentence. The results of Semantic role labeling in this paper solves a problem for labeling a single sentence semantic roles in Indonesian language, using computerized-based.

Keywords: *Concept-based Mining, Semantic Roles Labeling, Case Grammar*

Abstrak

Aplikasi pelabelan peran semantis sebagai salah satu *task* dari pemrosesan bahasa alami meliputi berbagai bidang, antara lain *machine translation*, *text summarization*, dan *text mining*. Pelabelan peran semantis digunakan antara lain pada *text mining* yaitu sebagai *preprocessing* dari *concept-based mining system*, yang memanfaatkan representasi data berbasis grafikal dan statistik untuk mencari konsep dari suatu kalimat dalam dokumen teks. Pencarian konsep pada *preprocessing* dari *text mining* merupakan salah satu cara untuk mendapatkan informasi yang berharga dari kalimat, dengan memperhatikan semantis setiap kata dalam kalimat. Hasil pelabelan peran semantis pada penelitian ini memecahkan masalah pelabelan peran semantis kalimat tunggal berbahasa Indonesia secara terkomputerisasi.

Kata Kunci: *Concept-based Mining, Pelabelan Peran Semantis, Tata Bahasa Kasus*

PENDAHULUAN

Latar Belakang

Text mining merupakan usaha untuk mendapatkan informasi yang berharga dari sekumpulan dokumen teks dengan memanfaatkan teknologi komputer, untuk mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen. *Text mining* juga merupakan proses penemuan akan informasi atau tren baru yang sebelumnya tidak terungkap dengan memproses dan menganalisis data dalam jumlah besar (Wibisono, 2006). *Text mining* sebagai upaya pengelolaan dokumen yang berupa teks, timbul dari kebutuhan akan pemahaman dan manipulasi dokumen teks untuk mendapatkan informasi tertentu dari sekumpulan dokumen teks dengan memanfaatkan teknologi komputer.

Sebagian besar teknik pengelompokan dokumen dalam *text mining* bekerja berdasarkan analisis dari frekuensi kemunculan kata atau frasa pada dokumen tanpa menghiraukan sisi semantis, sehingga ada kemungkinan dua kata atau frasa mempunyai frekuensi yang sama, padahal salah satu kata atau frasa tersebut memberikan kontribusi yang lebih besar dalam proses penentuan topik sebuah dokumen (Shehata, Karray & Kamel, 2007). Sebuah kata diabaikan maknanya dan juga kedudukannya dalam sebuah dokumen secara semantik, apakah berperan besar dalam mewakili isi dari suatu dokumen atau tidak. Hal ini tentu saja menyebabkan hasil yang diperoleh kurang akurat. Untuk memperoleh hasil yang lebih akurat, diperlukan pendekatan baru yang melibatkan peran semantis suatu kata dalam kalimat. Salah satu pendekatan baru tersebut adalah pemodelan berbasis konsep.

Pendekatan atau pemodelan berbasis konsep menangkap kata-kata yang merupakan konsep dari sebuah dokumen, yang nantinya dapat mewakili isi dokumen. Kedudukan konsep mempunyai peranan yang penting

dalam pengelompokan sebuah dokumen karena konsep-konsep tersebut merepresentasikan semantik dari dokumen dan semantik merepresentasikan arti dari dokumen. Dalam dokumen teks, sebuah kalimat yang lengkap dan berarti biasanya terdiri dari konsep utama yang memberikan gambaran tentang kandungan kalimat. Mendapatkan keterkaitan antara kata kerja dan argumen-argumen dalam sebuah kalimat mempunyai potensi yang menjanjikan dalam usaha untuk mengetahui arti dari sebuah kalimat. Informasi yang didapatkan tentang siapa sedang mengerjakan apa dan untuk siapa, dapat menjelaskan besarnya kontribusi setiap *term* (dapat berupa kata atau frasa) dalam suatu kalimat (Shehata, Karray & Kamel, 2007). Kata kerja dan argumen-argumennya merupakan konsep seperti yang telah disebutkan sebelumnya.

Pemodelan berbasis konsep terdiri atas 3 (tiga) bagian utama, yaitu *graphical concept-based*, *statistical concept-based*, dan *concept extractor* (Shehata, Karray & Kamel, 2007). Masing-masing bagian memiliki fungsi analisis dan peranan yang berbeda. *Graphical concept-based* menggambarkan struktur kalimat ke dalam bentuk graf konseptual dengan tetap mempertahankan semantik kalimat sesuai dengan dokumen aslinya. Setiap konsep pada *graphical concept-based* diberi bobot berdasarkan posisi konsep pada graf konseptual yang digambarkan. Sedangkan tujuan dari *statistical concept-based* adalah untuk memberikan bobot pada setiap konsep berdasarkan analisis dari frekuensi kemunculan konsep pada level kalimat dan dokumen. *Concept extractor* mengkombinasikan dua macam bobot yang dihasilkan oleh *graphical concept-based* dan *statistical concept-based* untuk dicari bobot konsep yang maksimal. Pembobotan konsep dilakukan menggunakan dua metode berbasis konsep, yaitu *graphical concept-based* di mana setiap konsep direpresentasikan ke dalam bentuk graf bertingkat dengan nilai yang berbeda untuk setiap levelnya dan *statistical concept-based* yang menganalisis konsep secara statistik

berdasarkan frekuensi kemunculannya. Konsep dengan nilai bobot tertinggilah yang akan mewakili topik suatu dokumen.

Masukan dari sistem berbasis konsep adalah kalimat yang telah melewati tahap *preprocessing*, yaitu kalimat yang telah diberi label peran semantis melalui tahap pelabelan peran semantis, yang dapat menunjukkan struktur *verb-argument*. Pelabelan peran semantis sebelum memasuki pemodelan berbasis konsep dilakukan untuk mendapatkan keterkaitan antara kata kerja dan argumen-argumennya dalam suatu kalimat atau dokumen teks.

Pelabelan peran semantis untuk kalimat berbahasa Indonesia secara lebih dalam akan dibahas dalam tulisan ini, di mana data yang digunakan untuk penelitian diambil dari artikel berita berbahasa Indonesia dan pelabelan peran semantis dilakukan berdasarkan teori tata bahasa kasus.

Ruang lingkup penelitian ini yaitu pelabelan peran semantis kalimat berbahasa Indonesia dengan menggunakan tata bahasa kasus. Pada penelitian ini, kalimat dibatasi pada kalimat deklaratif dan tunggal (bukan majemuk). Kalimat deklaratif adalah kalimat yang isinya menyatakan berita atau pernyataan untuk diketahui orang lain (pendengar atau pembaca). Kalimat berita dapat dibentuk dari satu atau lebih klausa, sesuai dengan besarnya atau luasnya isi berita yang ingin disampaikan. Contohnya : ABRI dan rakyat bahu-membahu membangun jembatan desa. Adapun pola kalimat yang digunakan pada penelitian ini terbatas pada pola kalimat S-P-O-K, S-P-O-Pel, S-P-O, S-P-Pel, S-P-K, dan S-P. Pola-pola kalimat tersebut dilihat berdasarkan ciri-ciri yang dimilikinya.

Data yang digunakan pada penelitian ini adalah sekumpulan artikel berita berbahasa Indonesia yang diambil dari *website* Harian Kompas (<http://www.kompas.com>) dan bersifat *offline*. Kalimat yang digunakan untuk pengujian dalam penelitian ini sebanyak 100 (seratus) kalimat, di mana 48 (empat puluh

delapan) kalimat diambil dari kalimat yang digunakan pada penelitian tata bahasa kasus untuk tanya-jawab (Musfirah, 2007) dan sisanya, 52 (lima puluh dua) kalimat, diambil dari Harian Kompas (<http://www.kompas.com>). Total artikel berita yang digunakan ada 300 artikel, yang diambil dari 6 (enam) topik dari Harian Kompas (<http://www.kompas.com>) yaitu internasional, pilkada, properti, bisnis dan keuangan, travel, dan regional, masing-masing sebanyak 50 (lima puluh) artikel. Sedangkan tipe semantik kata kerja didasarkan pada tipe semantik kata kerja yang digunakan pada penelitian tata bahasa kasus untuk tanya-jawab (Musfirah, 2007).

Tujuan Penelitian

Penelitian dilakukan untuk membangun sistem pelabelan peran semantis dengan menggunakan teknologi komputer, berdasarkan teori tata bahasa kasus. Dengan dibangunnya sistem pelabelan peran semantis ini, diharapkan dapat membantu manusia dalam pelabelan peran semantis, sehingga diperoleh keuntungan dari segi waktu (dengan menggunakan perangkat lunak, pelabelan peran semantis dapat dilakukan kapan saja, tidak terbatas pada jam kerja), geografis, kecepatan, dan biaya.

Pelabelan peran semantis sebagai bagian dari pemrosesan bahasa alami digunakan dalam pemrosesan dokumen karena pengguna menentukan relevansi dari dokumen dengan membaca dan menganalisisnya. Jika sistem pelabelan peran semantis dapat melakukan analisis dokumen secara otomatis, maka proses pencarian dokumen yang relevan akan lebih akurat. Misalnya, dalam penentuan topik, pengelompokan, dan pencarian dari sebuah dokumen sesuai dengan isi dari dokumen elektronik yang tersimpan dalam arsip yang relatif besar. Sebagian besar teknik pengelompokan dokumen dalam *text mining* bekerja berdasarkan analisis frekuensi kemunculan kata atau frasa pada dokumen tanpa menghiraukan sisi semantis. Hal ini

memungkinkan dua kata atau frasa mempunyai frekuensi yang sama, padahal salah satu kata atau frasa tersebut memberikan kontribusi yang lebih besar dalam proses penentuan topik sebuah dokumen.

Sebagai contoh, jika dilakukan pencarian surat atau dokumen elektronik pada aplikasi *e-office* dengan kata kunci pencarian sebagai berikut:

“Surat Tugas menghadiri Rapat Evaluasi”

Pemberian label berdasarkan peran semantis dilakukan terhadap argumen di mana “Surat Tugas” mendapatkan label subjek dan “Rapat Evaluasi Program Kerja” mendapat label objek, sehingga struktur *verb-argument* yang terbentuk dari kalimat tersebut adalah sebagai berikut:

“[ARG1 Surat Tugas] [TARGET menghadiri] [ARG2 Rapat Evaluasi]”

Hasil dari proses pelabelan secara semantis tersebut, yaitu verba dan argumen-argumennya kemudian digambarkan sebagai konsep beserta relasinya dalam representasi *concept-based mining model*, yaitu representasi *Conceptual Ontological Graph (COG)* dan dilakukan analisis terhadap masing-masing konsep pada level kalimat atau *ctf (conceptual term frequency)* dalam representasi *statistical concept-based mining model*.

Representasi *COG* menangkap konsep-konsep dan relasi di antara konsep-konsep tersebut (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Kemudian, konsep dan relasinya digambarkan ke dalam suatu representasi graf konseptual yang dibangun berdasarkan kalimat (Sheheta, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Setiap simpul pada representasi *COG* dapat berupa sebuah simpul konsep maupun sebuah graf konseptual. Jika sebuah simpul merujuk ke *graf* konseptual lainnya, hal ini berarti masih ada informasi yang lebih rinci mengenai topik yang ada pada simpul (konsep) tersebut, dan simpul tersebut digambarkan

sebagai konsep dan relasinya pada tingkat yang berada di atas graf konseptual yang dirujuknya (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Pada representasi *COG*, konsep-konsep dari kalimat diurutkan secara menurun (*descending*), yang berarti simpul tertinggi mewakili konsep yang paling umum dari kalimat dan simpul terendah mewakili konsep yang paling rinci dari kalimat (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Hal ini menghasilkan perbandingan konsep yang lebih informatif pada level kalimat dan level dokumen daripada jika hanya menggunakan perbandingan kata saja (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Representasi *COG* menyediakan beberapa macam level konsep yang bersarang dengan cara bertingkat (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Level ini dibangun berdasarkan tingkat kepentingan dari konsep dalam kalimat, yang akan digunakan untuk menganalisis konsep kunci dalam kalimat (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007). Representasi yang bertingkat dari *COG*, memberikan pemisahan antar konsep yang berkontribusi pada arti kalimat dengan jelas. Pemisahan ini diperlukan untuk membedakan antara konsep yang kurang berarti dengan konsep kunci pada kalimat. Konsep-konsep diletakkan pada representasi *COG* berdasarkan banyaknya *overlapping* antara istilah atau *term* (Sheheta, Karray & Kamel, 2005) (Sheheta, Karray & Kamel, 2006) (Sheheta, Karray & Kamel, 2007).

Pada representasi *statistical concept-based mining model*, *ctf* didefinisikan sebagai jumlah kemunculan konsep *c* dalam struktur *verb-argument* dari suatu kalimat *s* (Sheheta, Karray & Kamel, 2007). Untuk menganalisis masing-masing konsep pada level dokumen digunakan frekuensi *tf (term frequency)*, yaitu jumlah kemunculan sebuah konsep *c* dalam

dokumen asli dihitung (Sheheta, Karray & Kamel, 2007). Nilai *tf* dan *ctf* dihitung dengan algoritma *statistical analyzer*, yang menggambarkan proses penghitungan *tf* dan *ctf* konsep pada suatu dokumen (Sheheta, Karray dan Kamel, 2007). Algoritma dimulai dengan pemrosesan dokumen baru yang tersimpan dalam basis data atau gudang data, di mana kalimat-kalimatnya sudah dilabeli sesuai dengan peran semantis masing-masing kata dalam kalimat-kalimat tersebut (Sheheta, Karray & Kamel, 2007). Untuk setiap kalimat yang terlabeli, konsep pada struktur *verb-argument* yang merepresentasikan semantis kalimat diberi bobot sesuai nilai *ctf* dan *tf* (Sheheta, Karray & Kamel, 2007). Konsep yang bobotnya paling tinggi merupakan keluaran dari algoritma tersebut (Sheheta, Karray & Kamel, 2007). *Statistical concept-based mining model* merupakan faktor utama yang memperhatikan penting atau tidaknya suatu konsep pada sebuah dokumen dan pada kalimat (Sheheta, Karray & Kamel, 2007). Konsep dengan nilai tertinggi merupakan keluaran dari sistem.

Tinjauan Pustaka

Concept-Based Text Representation (Sheheta, Karray & Kamel, 2005)

Sebagian besar penelitian yang dilakukan terkait dengan *text mining* dalam kurun waktu 30 tahun didominasi oleh penggunaan metode statistik untuk mencocokkan antara bahasa alami yang digunakan oleh pengguna saat memasukkan *query* dengan data yang disimpan. Pendekatan *concept-based* merupakan solusi baru dalam *text mining*. Salah satu hal yang umum dalam riset *text mining* adalah dengan merepresentasikan teks sebagai kumpulan kata-kata atau lebih dikenal dengan istilah pendekatan *Bag-of-Words* (BoW). Kumpulan teks dalam suatu dokumen secara sederhana dianggap sebagai kumpulan kata-kata tanpa memperhatikan sintaksis yang merupakan salah satu cabang tata bahasa yang membicarakan struktur kalimat, klausa, dan frasa. Hal ini berarti

pada pendekatan BoW tidak mempedulikan urutan kata dalam dokumen karena kita hanya memperhatikan suatu kata sebagai entitas tersendiri yang tidak dipengaruhi oleh struktur kalimat di mana kata itu berada. Pendekatan BoW juga tidak memperhatikan unsur semantik yang merupakan cabang linguistik yang membahas tentang arti dan makna, yang artinya sebuah kata diabaikan maknanya dan juga kedudukannya dalam sebuah dokumen secara semantik apakah berperan besar dalam mewakili isi dari suatu dokumen atau tidak.

Untuk mengatasi kekurangan yang dialami oleh pendekatan BoW, maka muncul sebuah pendekatan baru yang dikenal dengan *Bag-of-Concepts* (BoC). Pendekatan BoC menangkap kata-kata yang merupakan konsep dari sebuah dokumen, yang nantinya dapat mewakili isi dokumen. Sebuah kalimat yang lengkap dan berarti biasanya terdiri dari konsep utama yang memberikan gambaran tentang kandungan kalimat. Mendapatkan keterkaitan antara *verb* dan argumen-argumennya dalam sebuah kalimat mempunyai potensi yang menjanjikan dalam usaha kita untuk mengetahui arti dari sebuah kalimat. Informasi yang kita dapatkan tentang siapa sedang mengerjakan apa dan untuk siapa dapat menjelaskan besarnya kontribusi setiap *term* dalam suatu kalimat. Verba dan argumen-argumennya merupakan konsep seperti yang telah disebutkan sebelumnya. Kedudukan konsep ini mempunyai peranan yang penting dalam pengelompokkan sebuah dokumen karena konsep-konsep tersebut merepresentasikan semantik dari dokumen dan semantik merepresentasikan arti dari dokumen.

Concept-Based Model (Sheheta, Karray, Kamal, 2005)

Pada *concept-based model*, pelabelan kalimat dilakukan berdasarkan peran semantis kata dalam kalimat, sehingga prosesnya disebut dengan *semantic role labeling*. Contoh sederhana *semantic role labeling* adalah sebagai berikut:

“Ibu membeli roti.”

Pada kalimat di atas terdapat sebuah verba yaitu membeli dan argumen-argumennya yaitu “Ibu” dan “roti”. Pemberian label berdasarkan peran semantis dilakukan terhadap argumen di mana “Ibu” mendapatkan label subjek dan “roti” mendapat label objek, sehingga struktur *verb-argument* yang terbentuk dari kalimat tersebut adalah sebagai berikut:

“[ARG0 Ibu] [TARGET membeli]
[ARG1 roti]”

Pada *concept-based model*, baik verba maupun argumen-argumennya dianggap sebagai *term*, berbeda dari metode *text mining* tradisional yang menganggap sebuah kata sebagai *term*. *Concept-based model* terdiri atas tiga bagian utama yaitu: *graphical concept-based*, *statistical concept-based*, dan *concept extractor*. *Graphical concept-based* menggambarkan struktur kalimat dengan tetap mempertahankan semantik kalimat sesuai dengan dokumen aslinya. Setiap konsep pada *graphical concept-based* diberi bobot berdasarkan posisi konsep pada struktur kalimat yang digambarkan. Sedangkan tujuan dari *statistical concept-based* adalah untuk memberikan bobot pada setiap konsep pada level kalimat dan dokumen. *Concept extractor* mengkombinasikan dua macam bobot yang dihasilkan oleh *graphical concept-based* dan *statistical concept-based* untuk dicari bobot konsep yang maksimal.

Pelabelan Peran Semantis

Pelabelan peran semantis timbul dari kebutuhan akan pemahaman dan pemanipulasian teks untuk mendapatkan informasi tertentu dari sekumpulan dokumen teks dengan memanfaatkan teknologi komputer namun tetap memperhatikan aspek sintaksis dan semantis, mengingat semakin besarnya volume berita elektronik berbahasa Indonesia yang di dalamnya terdapat berbagai informasi bermanfaat namun tidak bisa diperoleh secara langsung. Mengidentifikasi peran semantis dapat memberikan level analisis semantis yang bermanfaat dalam menyelesaikan *task*

pemrosesan bahasa alami. Peran semantis merepresentasikan partisipan dalam aksi atau keterhubungan yang digambarkan dengan kerangka semantis (Gildea dan Palmer, 2002).

Pelabelan peran semantis didefinisikan sebagai proses pengidentifikasian argumen dari predikat dalam suatu kalimat (Gildea dan Jurafsky, 2002). Ilustrasi dari pelabelan peran semantis, misalnya untuk kalimat “Ibu memotong roti.” adalah sebagai berikut. Pada kalimat “Ibu memotong roti.” terdapat sebuah kata kerja yaitu memotong dan argumen-argumennya yaitu “Ibu” dan “roti”. Pemberian label berdasarkan peran semantis dilakukan terhadap argumen di mana “Ibu” mendapatkan label subjek dan memiliki peran semantis sebagai pelaku (*agent*) serta “roti” mendapat label objek dan memiliki peran semantis sebagai penderita (*patient*), sehingga struktur *verb-argument* yang terbentuk dari kalimat tersebut adalah sebagai berikut:

“[ARG0 Ibu] [TARGET memotong]
[ARG1 roti]”

Label ARG0 dan ARG1, serta ARG3 sampai dengan ARG5 dan ARG-M mengacu pada label yang ada pada basis data PropBank (Gildea dan Jurafsky, 2002). PropBank merupakan salah satu implementasi dari pengembangan teori representasi peran semantis, dengan menambahkan lapisan semantis pada Penn TreeBank. Dalam PropBank, didefinisikan peran semantis mulai dari Arg0 hingga Arg5 serta ArgM-X untuk argumen yang spesifik. Misalnya, ArgM-LOC untuk argumen yang menerangkan lokasi, dan ArgM-TEMP untuk argumen waktu (Palmer, Kingsbury dan Gildea, 2002).

PropBank merupakan salah satu basis data semantis yang digunakan untuk pelabelan peran semantis kalimat berbahasa Inggris. Tujuan pembangunan basis data PropBank adalah menyediakan *hand-annotated corpus* dan memungkinkan pembangunan *domain-independent* dari sistem pemahaman bahasa. Pada basis data PropBank, didefinisikan sekumpulan peran semantis untuk setiap kata

kerja, dan memberikan keterangan untuk setiap kasus pada Penn TreeBank yang asli. Setiap peran dari kata kerja diberikan nomor, sebagai contoh, pada kata kerja “offer”:

...[Arg0 the company] to ... [Arg1 a 15% stake] [Arg2 to the public]. (wsj_0345)

...[Arg0 Sotheby's] ... offered [Arg2 the Dorrance heirs] [Arg1 a money-back guarantee] (wsj_1928)

...[Arg1 an amendment] offered [Arg0 by Rep.Peter DeFazio] ... (wsj_0107)

...[Arg2 Subcontractors] will be offered [Arg1 a settlement] ... (wsj_0187)

Dengan melabeli peran semantis untuk setiap kata kerja dalam *corpus*, PropBank menyediakan sumber yang sifatnya *domain-independent*, dengan harapan dapat menghasilkan sistem pemrosesan bahasa alami yang luas cakupannya. Fokus dari *PropBank* adalah pada struktur argumen dari kata kerja, dan menyediakan *annotated-corpus* dengan peran semantis, termasuk peran tradisional yang ditampilkan sebagai argumen dan sebagai keterangan. Pembangunan basis data PropBank untuk bahasa Inggris memerlukan waktu sekitar 3 (tiga) tahun, dengan melibatkan sekitar 30 (tiga puluh) orang ahli bahasa. Selain ahli bahasa, juga ahli-ahli dari bidang lain serta mahasiswa dari berbagai latar belakang ilmu.

Secara umum, pelabelan peran semantis dengan memanfaatkan basis data PropBank terbagi menjadi dua langkah, yaitu pengidentifikasian elemen kerangka dan pelabelan setiap kalimat dengan elemen kerangka yang sudah teridentifikasi. Pelabelan peran semantis menghasilkan kata atau frasa yang telah dilabeli sesuai peran semantisnya dalam kalimat. Kata atau frasa yang telah dilabeli inilah yang disebut dengan konsep. Pencarian konsep pada tahap *preprocessing* dari sistem berbasis konsep merupakan salah satu cara untuk mendapatkan informasi

yang berharga dari suatu kalimat dengan memperhatikan makna dari setiap kata dalam suatu kalimat.

Pada penelitian ini, pelabelan peran semantis dilakukan berdasarkan teori tata bahasa kasus, di mana untuk setiap kalimat dicari kata kerja dan argumennya sehingga setiap kata kerja mempunyai struktur *verb-argument* masing-masing, sedangkan labelnya diambil dari PropBank.

Tata Bahasa Kasus untuk Pelabelan Peran Semantis

Salah satu teori yang dapat menyelesaikan permasalahan pelabelan peran semantis adalah tata bahasa kasus. Tata bahasa kasus merupakan salah satu teori mengenai representasi peran semantis. Tata bahasa kasus dapat mendeskripsikan relasi semantis yang terdapat dalam suatu kalimat (Albright, 2006).

Tata bahasa kasus merupakan suatu pendekatan terhadap tata bahasa yang memberi penekanan pada hubungan-hubungan semantis dalam suatu kalimat. Dalam tata bahasa kasus, kata kerja dianggap sebagai bagian kalimat yang paling penting, dan memiliki sejumlah hubungan semantis dengan satu atau lebih frasa nomina (kata benda). Hubungan-hubungan inilah yang disebut kasus. Kata kerja dipandang sebagai suatu peristiwa dan kasus merupakan peran frasa nomina pada peristiwa tersebut (Tarigan, 1990).

Struktur semantis terdiri dari dua unit semantis pokok, yaitu kata kerja dan kata benda. Dalam struktur semantis ini kata kerja merupakan pusat. Ini berarti bahwa kata kerja menentukan hadirnya kata benda dalam struktur semantis. Kata kerja dimaksud bukan hanya mencakup pengertian kata kerja yang dikenal pada struktur luar secara tradisional, seperti makan, tidur, dan lain-lain, tetapi juga kata sifat dan kata-kata benda. Dengan kata lain, semua kata yang berfungsi sebagai predikat dalam kalimat dianggap sebagai kata kerja dalam struktur semantisnya. (Tarigan, 1990).

Relasi semantis dalam tata bahasa kasus dideskripsikan sebagai relasi antar kasus. Kasus dalam tata bahasa kasus sendiri dapat mendeskripsikan relasi antara kata kerja dan kata benda atau argumennya. Struktur yang berkaitan dengan makna dari kata kerja dideskripsikan dalam kerangka kasus. Kata kerja akan diikuti dengan sekumpulan kasus yang mungkin, yang berkaitan dengannya.

Secara umum, pelabelan peran semantis berdasarkan tata bahasa kasus dilakukan dalam dua langkah. *Pertama*, mengidentifikasi elemen kerangka. Setelah diketahui struktur sintaksis dari suatu kalimat dalam sebuah dokumen teks, maka selanjutnya berdasarkan struktur sintaksis tersebut akan dicari kerangka kasus dari kata kerja yang terdapat dalam kalimat tersebut. *Kedua*, melabeli setiap kalimat dengan elemen kerangka yang sudah teridentifikasi. Setelah kerangka kasus pada langkah pertama ditemukan, maka langkah berikutnya adalah menentukan peran semantis dari setiap kata dalam kalimat. Setelah setiap kata dalam kalimat diketahui peran semantisnya, kemudian dilakukan pelabelan peran semantis untuk setiap kata yang telah ditentukan atau diketahui peran semantisnya. Hasil pelabelan peran semantis tersebut kemudian disimpan dalam *file* tertentu (pada penelitian ini, *file* yang digunakan untuk menyimpan berupa *file* berekstensi .txt) dan kemudian digunakan dalam proses berikutnya pada *text mining*, yaitu sebagai masukan pada pemodelan berbasis konsep.

Lima Kasus Proposisi

Struktur dalam kalimat terdiri dari suatu modalitas yang berisi atau mengandung frasa dan suatu preposisi yang berisi atau mengandung suatu kata kerja ditambah satu atau lebih kategori-kategori kasus yang berbeda. Suatu struktur proposisi kasus terdiri dari satu unsur kata kerja sebagai pusat dan yang disebut juga predikat, dan satu unsur kata benda atau lebih yang bergantung pada kata kerja itu (Tarigan, 1990).

Kasus proposisi adalah kasus yang merupakan valensi kata kerja. Dengan kata lain, kehadirannya dalam struktur semantis ditentukan oleh kata kerja. Ahli tata bahasa kasus Fillmore dan Cook mengemukakan, suatu struktur proposisi kasus terdiri satu unsur kata kerja sebagai pusat dan satu atau lebih satu unsur kata benda yang bergantung pada kata kerja itu, unsur kata benda ini disebut kasus. Ada lima kasus proposisi (Tarigan, 1990), yaitu *Agent* (Agen), *Experiencer* (Pengalam), *Benefactive* (Benefaktif), *Object* (Objek), dan *Locative* (Lokatif).

Agen (A) menyatakan suatu *entiti* (*entity*) yang menjadi penuar (*instigator*) suatu aksi yang dinyatakan oleh kata kerja (Tarigan, 1990). *Entiti* yang dimaksud pada dasarnya harus hidup atau bernyawa (Tarigan, 1990). Pengalam(P) menyatakan *entiti* yang mengalami sesuatu secara kognisi, emosi, atau sensasi (Tarigan, 1990). *Entiti* yang dimaksud pada dasarnya harus hidup atau bernyawa (Tarigan, 1990). Benefaktif (B) menyatakan *entiti* yang memiliki, mendapat, atau kehilangan sesuatu (Tarigan, 1990). *Entiti* yang dimaksud harus hidup atau bernyawa. Objek (O) menyatakan *entiti* yang berada dalam suatu keadaan, *entiti* yang berada dalam suatu proses, *entiti* yang terkena pengaruh suatu aksi atau merupakan hasil suatu aksi, Objek (O) juga menyatakan *entiti* yang merupakan suatu pengalaman yang bersifat kognisi, emosi, atau sensasi (Tarigan, 1990). Selain itu juga menyatakan *entiti* yang berada dalam pemilikan, yang hilang, atau yang berpindah lokasi (Tarigan, 1990).

Tipe Kata Kerja Dasar

Dalam penelitian mengenai tata bahasa kasus untuk tanya-jawab (Musfirah, 2007) disebutkan bahwa berdasarkan penelitian yang dilakukan oleh Anwar Ibrahim, mahasiswa Fakultas Keguruan Sastra dan Seni IKIP Medan, dilihat dari ciri-ciri semantisnya terdapat dua belas tipe kata kerja dasar. Dua belas tipe kata kerja dasar beserta kasus yang diperlukan dapat dilihat pada Tabel 1.

Tabel 1. Dua Belas Tipe Kata Kerja Dasar

No.	Tipe Kata Kerja Dasar	Kasus yang Diperlukan	Kerangka Kasus
	KK Keadaan	Objek keadaan	+ [- Ok]
	KK Keadaan – Pengalam	Pengalam, Objek keadaan	+ [- P, Ok]
	KK Keadaan – Benefaktif	Benefaktif, Objek keadaan	+ [- B, Ok]
	KK Keadaan – Lokatif	Objek keadaan, Lokatif	+ [- Ok, L]
	KK Proses	Objek	+ [- O]
	KK Proses – Pengalam	Pengalam, Objek	+ [- P, O]
	KK Proses – Benefaktif	Benefaktif, Objek	+ [- B, O]
	KK Proses – Lokatif	Objek, Lokatif	+ [- O, L]
	KK Aksi	Agen, Objek	+ [- A, O]
	KK Aksi – Pengalam	Agen, Pengalam, Objek	+ [- A, P, O]
	KK Aksi – Benefaktif	Agen, Benefaktif, Objek	+ [- A, B, O]
	KK Aksi – Lokatif	Agen, Objek, Lokatif	+ [- A, O, L]

Urutan kasus setiap KK disebut kerangka kasus (*case frame*) KK itu, dan secara konvensional dituliskan: + [- Ok], + [- P, Ok], + [- B, Ok], + [- Ok, L], dan seterusnya. Tanda garis (-) dalam kerangka kasus tersebut menandakan bahwa ada KK tertentu yang dapat dimasukkan dalam kerangka kasus bersangkutan. Tanda “+” menyatakan ciri semantis. Urutan kasus setiap KK disebut kerangka kasus (*case frame*) KK itu, dan secara konvensional dituliskan: + [- Ok], + [- P, Ok], + [- B, Ok], + [- Ok, L], dan seterusnya. Tanda garis (-) dalam kerangka kasus tersebut menandakan bahwa ada KK tertentu yang dapat dimasukkan dalam kerangka kasus bersangkutan. Tanda “+” menyatakan ciri semantis.

Kasus atau elemen kerangka yang diperlukan pada tipe kata kerja keadaan adalah Objek keadaan (Ok), sehingga kerangka kasusnya adalah + [- Ok]. Kasus yang diperlukan pada tipe kata kerja keadaan-pengalam adalah Pengalam (P) dan Objek keadaan (Ok), sehingga kerangka kasusnya adalah + [- P, Ok]. Kasus yang diperlukan pada tipe kata kerja keadaan-benefaktif adalah Benefaktif (B) dan Objek keadaan (Ok), sehingga kerangka kasusnya adalah + [- B, Ok]. Kasus yang diperlukan pada tipe kata kerja keadaan-lokatif adalah Objek keadaan (Ok)

dan Lokatif (L), sehingga kerangka kasusnya adalah + [- Ok, L]. Kasus yang diperlukan pada tipe kata kerja proses adalah Objek (O), sehingga kerangka kasusnya adalah + [- O]. Kasus yang diperlukan pada tipe kata kerja proses-pengalam adalah Pengalam (P) dan Objek (O), sehingga kerangka kasusnya adalah + [- P, O]. Kasus yang diperlukan pada tipe kata kerja proses-benefaktif adalah Benefaktif (B) dan Objek (O), sehingga kerangka kasusnya adalah + [- B, O]. Kasus yang diperlukan pada tipe kata kerja proses-lokatif adalah Objek (O) dan Lokatif (L), sehingga kerangka kasusnya adalah + [- O, L]. Kasus yang diperlukan pada tipe kata kerja aksi adalah Agen (A) dan Objek (O), sehingga kerangka kasusnya adalah + [- A, O]. Kasus yang diperlukan pada tipe kata kerja aksi-pengalam adalah Agen (A), Pengalam (P), dan Objek (O), sehingga kerangka kasusnya adalah + [- A, P, O]. Kasus yang diperlukan pada tipe kata kerja aksi-benefaktif adalah Agen (A), Benefaktif (B) dan Objek (O), sehingga kerangka kasusnya adalah + [- A, B, O]. Kasus yang diperlukan pada tipe kata kerja aksi-lokatif adalah Agen (A), Objek (O), dan Lokatif (L), sehingga kerangka kasusnya adalah + [- A, O, L]. Jika telah ditemukan kerangka kasusnya, maka akan dibangkitkan label peran semantis sesuai dengan label yang ada pada basis data PropBank.

Penelitian Sebelumnya

Penelitian mengenai pelabelan peran semantis telah dilakukan oleh para ahli, salah satunya untuk dokumen teks berbahasa Inggris. Pada pelabelan peran semantis untuk dokumen teks berbahasa Inggris, digunakan *classifier* statistik untuk mendukung pelabelan peran semantis kalimat berbahasa Inggris secara otomatis. Penelitian-penelitian yang dilakukan untuk pelabelan peran semantis pada umumnya dilakukan berdasarkan basis data PropBank (Palmer, Kingsbury dan Gildea, 2002) dan FrameNet (Palmer, Kingsbury dan Gildea, 2002) untuk berbagai bahasa, antara lain untuk bahasa Swedia dan Basque.

Pada penelitian pelabelan peran semantis untuk bahasa Indonesia ini, digunakan tata bahasa kasus untuk menyelesaikan masalah pelabelan peran semantis. Basis data *PropBank* untuk bahasa Indonesia tidak dibangun tetapi digunakan sebagai pedoman dalam pelabelan peran semantis. Hasil dari pelabelan peran semantis akan dijadikan dasar untuk mencari konsep dari suatu kalimat. Setelah itu dilakukan analisis grafikal dan statistik, di mana kalimat akan direpresentasikan dalam bentuk graf dan untuk masing-masing konsep akan diberikan bobot atau nilai tertentu.

Theoretical Framework

Secara umum, proses pelabelan peran semantis berdasarkan tata bahasa kasus dilakukan dengan langkah-langkah sebagai berikut:

1. Identifikasi kerangka kasus
2. Melabeli setiap kalimat dengan elemen kerangka yang sudah teridentifikasi

Metode Penelitian

Perumusan masalah pada penelitian ini dilakukan dengan cara studi literatur. Sedangkan data penelitian ini didapatkan melalui pengumpulan “kalimat dari media

elektronik”, yang dikumpulkan dalam bentuk dokumen teks. Sedangkan pelabelan peran semantis dilakukan sesuai dengan teori tata bahasa kasus. Pembangunan sistem aplikasi pelabelan peran semantis dilakukan dengan cara analisis kebutuhan, perancangan sistem sesuai dengan hasil analisis kebutuhan, implementasi, dan pengujian sistem yang sudah dibangun.

Pada penelitian ini, digunakan tata bahasa kasus untuk menyelesaikan masalah pelabelan peran semantis. Sedangkan data yang digunakan untuk pengujian artikel berita berbahasa Indonesia, yang diambil dari web dan bersifat *offline*. Sebelum dilakukan pelabelan peran semantis, dilakukan penguraian sintaksis dari masing-masing kalimat untuk mengetahui peran sintaksis dari setiap kata dalam suatu kalimat.

Langkah pertama dari pelabelan peran semantis adalah mencari kata kerja dari suatu kalimat tunggal berbahasa Indonesia. Setelah ditemukan kata kerjanya, akan dilakukan identifikasi kerangka kasus. Setelah ditemukan kerangka kasus yang sesuai, akan ditentukan peran semantis dari setiap kata dalam suatu kalimat, di mana peran kata-kata tersebut direpresentasikan dalam tata bahasa kasus. Kemudian dibangkitkan label peran semantis berdasarkan PropBank untuk setiap kata dalam kalimat tersebut. Kalimat yang telah diberi label peran semantis kemudian disimpan dalam sebuah *file*, sehingga dapat digunakan untuk proses berikutnya. Misalnya untuk *text mining*, di mana digunakan *concept-based mining model* atau pemodelan berbasis konsep dalam klasifikasi dan kategorisasi dokumen teks. Berdasarkan hasil pelabelan peran semantis akan dicari konsep dari suatu kalimat. Setelah itu dilakukan analisis grafikal dan statistik, di mana kalimat akan direpresentasikan dalam bentuk graf dan untuk masing-masing konsep akan diberikan bobot atau nilai tertentu. Kedua analisis tersebut dapat memperbaiki performansi dari klasifikasi dan kategorisasi dokumen teks.

HASIL DAN PEMBAHASAN

Pengurai Sintaksis

Penguraian sintaksis menunjukkan analisis sintaks pada kalimat berdasarkan pada teori tata bahasa. Informasi-informasi yang diperoleh dari proses penguraian sintaksis diperlukan dalam pengidentifikasian argumen kalimat. Identifikasi frasa kata benda yang kompleks dan menentukan bahwa frasa tersebut merupakan argumen dari suatu kata kerja tertentu dalam proses pelabelan peran semantis akan sulit dilakukan jika tidak ada informasi dari pengurai sintaksis.

Tool pengurai sintaksis yang digunakan pada penelitian ini adalah PC-PATR (Musfirah, 2007). PC-PATR merupakan implementasi dari metode *constraint-based formalisms* pada *personal computer (PC)*. PC-PATR ini dapat dijalankan pada beberapa sistem operasi seperti misalnya : MS-DOS, *Microsoft Windows*, Macintosh, dan Unix. *Constraint-based formalism* yang digunakan pada PC-PATR disebut dengan istilah PATR-II *formalism* (Musfirah, 2007).

Grammar PATR-II terdiri dari himpunan *rule* (aturan) dan *lexicon* (kamus kata). Setiap *rule* terdiri dari *phrase structure rule* dan himpunan *feature constraint*. *Feature constraint* adalah *unification* pada *feature structure* yang berhubungan dengan konstituen dari *phrase structure rule*. Sedangkan *lexicon* menyediakan *record-record* yang menggantikan simbol terminal pada *phrase structure rule*. Setiap *record* ini berisi kata beserta *feature*-nya. (Musfirah, 2007).

PC-PATR (Musfirah, 2007) dijalankan dengan terlebih dahulu membaca (*load*) *file grammar* dan *file lexicon*-nya. Kemudian PC-PATR akan membaca kalimat masukan, menguraikannya berdasarkan *grammar* dan *lexicon*, lalu menghasilkan keluaran berupa pohon urai kalimat tersebut. Kalimat-kalimat masukan ini bisa dituliskan di dalam *file* berekstensi .sen. Hasil keluarannya dapat disimpan dalam *file* berekstensi .log.

File grammar adalah *file* yang berisi *phrase structure rule*, termasuk *feature constraint*-nya. Sebelum definisi *rule-rule*, pada *file grammar* dapat dibuat suatu *feature template*. *Feature template* digunakan sebagai makro di *lexicon*. *Feature template* dapat digunakan untuk memberikan nilai *default feature structure* terhadap suatu kategori. Contoh: Let N be <subcat> = !istilah, artinya kategori N didefinisikan memiliki subcat yang nilai default-nya adalah istilah. Efek dari *template* ini adalah membuat semua nomina (N) mempunyai subkategori istilah, kecuali didefinisikan lain. Jika tidak ada tanda seru (!) maka semua nomina didefinisikan mempunyai subkategori istilah, tanpa terkecuali.

Lexicon merupakan *file* basis data bentuk standar yang memiliki sejumlah *record*. Masing-masing *record* merepresentasikan kata. Setiap *record* dibagi atas beberapa *field*, di mana masing-masing *field* didahului oleh suatu penanda bentuk standar pada awal barisnya. Penanda ini didahului oleh karakter “\” dan diikuti oleh satu atau lebih karakter alfanumerik. Penanda yang digunakan dalam penelitian ini adalah:

\w, adalah penanda untuk bentuk leksikal kata, di mana pengejaannya sama persis dengan pengejaan kata yang akan dijadikan masukan PC-PATR.

\c, adalah penanda untuk kategori/kelas kata

\f, adalah penanda untuk ciri tambahan dari kata

Contoh isi *file lexicon* untuk kata menyiapkan:

\w menyiapkan

\c Vtran

\f <subcat> = ekatransitif

File sentence berisi kalimat masukan yang hendak diuraikan. Satu kalimat terletak di satu baris. Jika kalimat yang ingin diuraikan ada lebih dari satu, maka kalimat-kalimat tersebut harus disusun baris per baris. Kalimat yang sangat panjang tidak boleh dipisah menjadi 2 (dua) baris atau lebih, karena PC-

PATR menguraikan *file sentence* ini baris demi baris. *File .log* adalah *file* keluaran. *File* ini berisi hasil penguraian *file sentence* terhadap *grammar* dan *lexicon* yang digunakan.

Pada penelitian ini, data yang telah diperoleh dari Harian Kompas (<http://www.kompas.com>) dipindahkan ke dalam *file* berekstensi .sen dan kemudian dilakukan penguraian dengan menggunakan pengurai sintaksis. Hasil penguraian kemudian disimpan pada *file* berekstensi .log dan diproses oleh sistem pelabelan peran semantis. Selain memberikan keterangan mengenai peran sintaksis dari setiap kata dalam kalimat, hasil penguraian sintaksis tersebut dapat memberikan informasi mengenai tipe dari setiap kata dalam kalimat, yang dalam bahasa Inggris biasa disebut sebagai *part-of speech* (POS).

Pola Kalimat

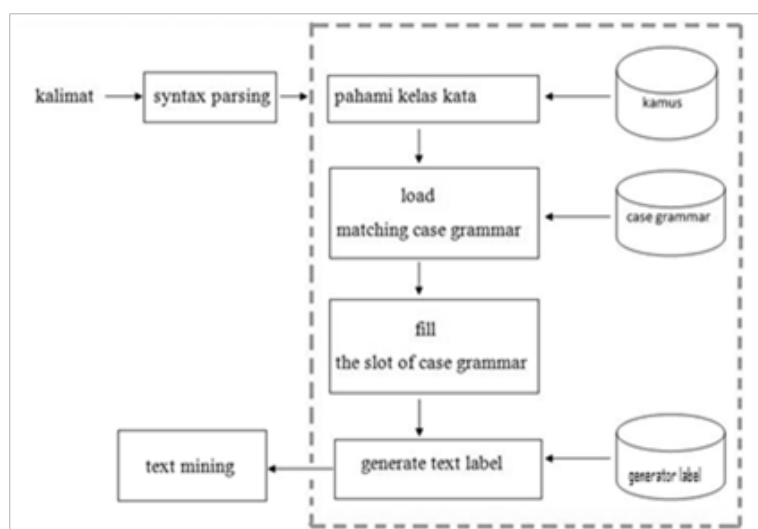
Kalimat tunggal adalah kalimat yang terdiri atas satu klausa. Hal tersebut berarti bahwa konstituen untuk tiap unsur kalimat, seperti subjek dan predikat, hanyalah satu atau merupakan satu kesatuan. Dalam kalimat tunggal terdapat semua unsur wajib yang diperlukan tetapi tidak menutup kemungkinan ada juga unsur lain seperti keterangan tempat, keterangan waktu, dan keterangan alat.

Kalimat berpola S-P-O-K mempunyai unsur subjek, predikat, objek, dan keterangan. Subjek dapat berupa nomina atau frasa nominal, predikat berupa verba dwitransitif, objek berupa nomina atau frasa nominal, dan keterangan berupa frasa berpreposisi. Kalimat berpola S-P-O-Pel adalah kalimat yang mempunyai unsur subjek, predikat, objek, dan pelengkap; subjek berupa nomina atau frasa nominal, predikat berupa kata kerja dwitransitif, objek berupa nomina atau

frasa nominal, dan pelengkap berupa nomina atau frasa nominal. Kalimat berpola S-P-O mempunyai unsur subjek, predikat, dan objek; subjek berupa nomina atau frasa nominal, predikat berupa verba transitif, dan objek berupa nomina atau frasa nominal. Kalimat berpola S-P-K mempunyai unsur subjek, predikat, dan harus memiliki unsur keterangan karena diperlukan oleh predikat. Subjek berupa nomina atau frasa nominal, predikat berupa verba intransitif, dan keterangan berupa frasa berpreposisi. Sedangkan kalimat berpola S-P merupakan kalimat dasar yang mempunyai unsur subjek dan predikat.

Perancangan Sistem Pelabelan Peran Semantis

Masukan dari sistem pelabelan peran semantis adalah kalimat tunggal berbahasa Indonesia yang telah diuraikan secara sintaksis. Sistem akan memahami dan menentukan kelas kata dari kata kerja dalam kalimat yang telah diuraikan tadi. Proses berikutnya adalah mendapatkan kerangka kasus yang sesuai dengan kata kerja, dari basis data. Basis pengetahuan tata bahasa kasus berisi aturan untuk menentukan peran semantis kalimat tersebut. Setelah ditemukan kerangka kasus yang sesuai, akan dilakukan pengisian *slot* dari tata bahasa kasus tersebut. Terakhir, akan dibangkitkan label berdasarkan peran semantis masing-masing kata dalam kalimat. Sebagai hasil akhir dari perangkat lunak ini adalah kalimat yang sudah diberi label peran semantis. Kalimat yang telah diberi label ini kemudian disimpan dalam *file* berekstensi .txt, untuk kemudian digunakan sebagai masukan dari sistem pemodelan berbasis konsep (Shehata, Karray dan Kamel, 2007). Jika digambarkan, secara umum prosesnya seperti pada Gambar 1.



Gambar 1. Gambaran Umum Sistem Pelabelan Peran Semantis

Dari Gambar 1 dapat dilihat bagian yang berada dalam garis putus-putus merupakan sistem yang dibangun dalam penelitian ini. Dalam sistem pelabelan peran semantis yang dibangun berdasarkan teori tata bahasa kasus (*case grammar*) pada penelitian ini, masukan ke sistem adalah sebuah kalimat tunggal yang diambil dari artikel berita yang berasal dari *website* media cetak yang diambil secara *offline*. Kalimat yang diambil dari artikel berita tersebut akan dimasukkan ke dalam *file* berekstensi *.sen*. Kemudian kalimat tersebut akan di-*parsing* dengan menggunakan PC-PATR, kemudian disimpan dalam *file* berekstensi *.log*. *File .log* berisi pohon urai yang menunjukkan struktur sintaksis suatu kalimat. Untuk mendapatkan pohon urai tersebut, kalimat yang akan diproses atau dicari pohon urainya disimpan dalam *file* berekstensi *.log*. Kemudian dijalankan *parser* PC-PATR untuk mencari pohon urai dari kalimat yang telah disimpan dalam *.sen* tadi, dengan perintah *console*, yang secara otomatis akan menyimpan pohon urai hasil *parsing* ke dalam *file .log* (nama *file .log* dapat ditentukan oleh pengguna). Berdasarkan hasil *syntax parsing* dengan menggunakan PC-PATR, sistem akan memahami dan menentukan kelas kata dari kata kerja dalam kalimat yang telah di-*parsing* tadi. Proses berikutnya adalah *load*

case grammar yang sesuai dengan kata kerja, dari basis data. Basis pengetahuan tata bahasa kasus di sini berisi aturan untuk menentukan peran semantis kalimat tersebut. Setelah ditemukan *case frame* yang sesuai, akan dilakukan pengisian *slot* dari *case grammar* tersebut. Terakhir, akan di-*generate* label berdasarkan peran semantis masing-masing kata dalam kalimat. Sebagai hasil akhir dari perangkat lunak ini adalah kalimat yang sudah diberi label peran semantis. Kalimat yang telah diberi label ini kemudian disimpan dalam *file* berekstensi *.txt*, untuk kemudian digunakan sebagai masukan dari sistem *concept-based model* (Shehata, Karray dan Kamel, 2007).

Kebutuhan fungsional sistem *semantic role labeling* yang dibangun dalam penelitian ini, yaitu: menentukan fungsi sintaksis, menentukan label peran semantis, kelola kamus semantik, kelola kamus kata kerja, kelola basis pengetahuan *case grammar*, kelola *generator*. Basis data kata kerja diambil dari hasil penelitian mahasiswa Fakultas Sastra IKIP Medan, yang digunakan dalam penelitian “Penentuan Peran Semantis Kata dalam Kalimat Bahasa Indonesia Menggunakan *Case Grammar*” (Sawitri, 2007). Pada sistem pelabelan peran semantis ini, *generator* label merupakan aturan pelabelan peran semantis.

Aturan-aturan atau *rule-rule* ini disimpan dalam Tabel Generatorlabel. Pendekatan yang dilakukan ialah membangun daftar *generator* label, di mana ada 3 (tiga) hal, yaitu:

a. Pola kalimat deklaratif

Tipe label merupakan peran semantis dari setiap kata dan setiap kata kata memiliki label yang sesuai.

b. *Case grammar*

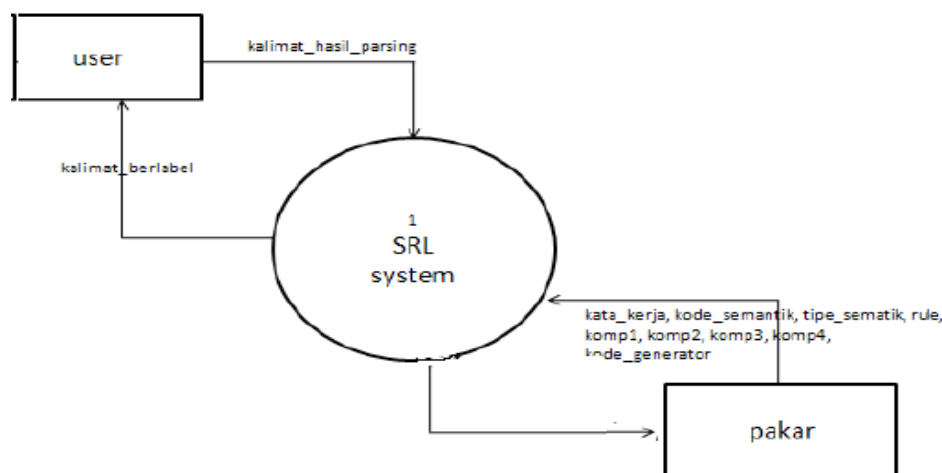
Pada *case grammar* terdapat dua belas tipe struktur semantis kalimat deklaratif. Struktur tersebut mendeskripsikan satu kata kerja beserta kasus-kasus yang harus muncul berdasarkan tipe semantik kata kerjanya. Kata kerja dipandang sebagai sebuah *event*, dan kasus-kasus merupakan

peran semantis unsur-unsur kata benda terhadap sebuah *event*. Pembentukan *generator* label memanfaatkan dua belas tipe struktur semantis *case grammar*.

c. Label peran semantis dari PropBank

PropBank mendefinisikan label peran semantis mulai dari ARG-0 hingga ARG-5 untuk label non-keterangan, dan ARG-M untuk memberikan label pada kata-kata yang secara sintaksis memiliki peran sebagai keterangan. Label ARG0 diberikan pada setiap kata yang memiliki peran semantis sebagai agen.

Diagram konteks dan diagram aliran data dari sistem pelabelan peran semantis dapat dilihat pada Gambar 2 dan Gambar 3 berikut.



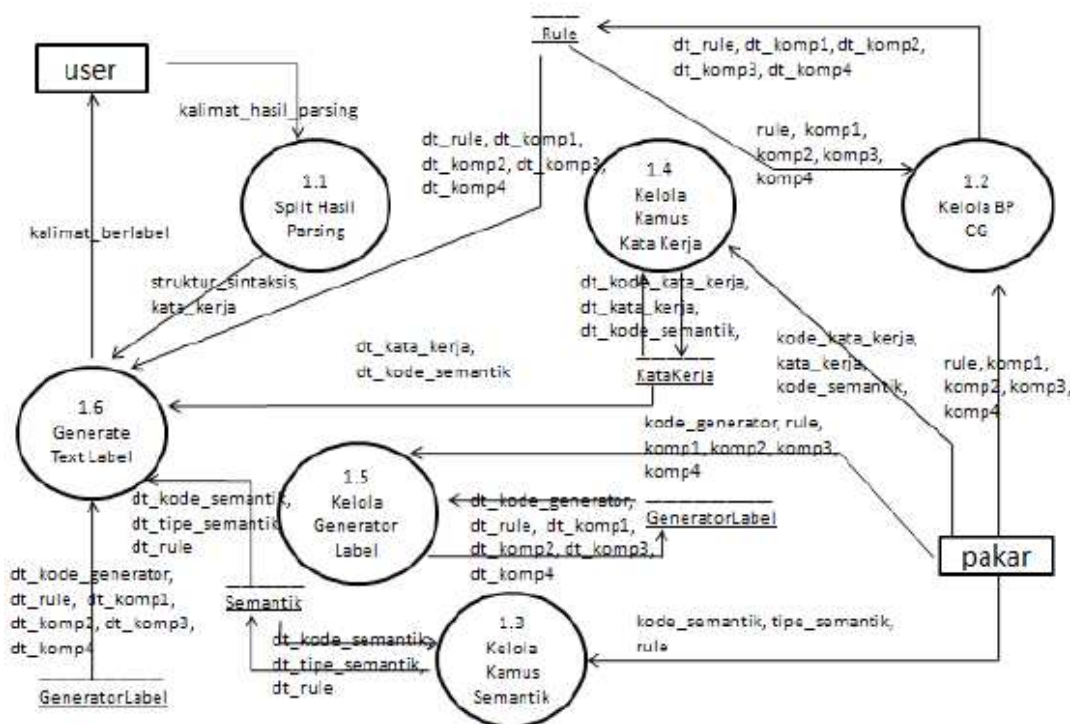
Gambar 2. Diagram Konteks Sistem Pelabelan Peran Semantis (SRL System)

Pada Gambar 2 dapat dilihat, *user* (pengguna) memberikan *input* berupa kalimat hasil *parsing* ke sistem pelabelan peran semantis (*SRL system*). Sistem akan mengolah masukan dan memberikan keluaran berupa

kalimat yang sudah diberikan label peran semantis. Proses pelabelan peran semantis secara lebih detail dapat dilihat pada Gambar 3 (diagram aliran data (DAD) sistem pelabelan peran semantis (*SRL system*)).

Pada Gambar 3 dapat dilihat diagram aliran data (DAD) dari sistem pelabelan peran semantis (SRL system). Sistem pelabelan peran semantis (SRL system) melakukan fungsi *split*

hasil *parsing*, kelola tata bahasa kasus, kelola kamus semantis, kelola kamus kata kerja, kelola generator label, dan *generate text label*.



Gambar 3. Diagram Aliran data (DAD) Sistem Pelabelan Peran Semantis (SRL System)

Implementasi Pelabelan Peran Semantis

Dalam penelitian ini dibuat suatu perangkat lunak yang memiliki fungsi utama sebagai aplikasi *semantic role labeling* dengan menggunakan *case grammar* yang telah disebutkan sebelumnya. Inputan dari perangkat lunak ini adalah sekumpulan *file* berupa *file .log*. Hasil dari sistem ini adalah kalimat yang sudah berlabel, sesuai dengan teori dan metode yang telah disebutkan sebelumnya. Pada intinya, input atau masukan sistem adalah *file* hasil *syntax parsing* dengan menggunakan *parser* PC-PATR. *File-file* hasil *parsing* ini berbentuk *file .log*. Sedangkan keluaran dari sistem adalah kalimat yang sudah dibari label peran semantis sesuai dengan anotasi *PropBank*.

Implementasi dalam penelitian ini menggunakan perangkat lunak sebagai berikut:

- Sistem Operasi : Windows Vista Ultimate
- Bahasa Pemrograman : PHP
- Basis Data : MySQL

Adapun spesifikasi perangkat keras yang digunakan dalam pembangunan sistem adalah sebagai berikut:

- Processor Intel Pentium Dual CPU 1,73 GHz
- Memori (RAM) 1014 MB
- Hardisk 200 GB

Data set yang digunakan dalam penelitian ini adalah kumpulan artikel-artikel yang berasal dari media surat kabar Kompas

(<http://www.kompas.com>) yang terbagi dalam topik sebagai berikut (mengacu pada topik yang ada dalam *Harian Kompas Online*), yaitu: bisnis dan keuangan, kesehatan, properti, regional, travel, dan internasional. Tabel-tabel diimplementasikan dengan menggunakan MySQL, sesuai dengan hasil perancangan. Basis data diimplementasikan dengan nama SRL_database, dengan daftar tabelnya dapat dilihat pada Tabel 2 berikut:

Tabel 2. Implementasi Basis Data

No	Nama Tabel	Deskripsi Tabel
1	Semantik	Menyimpan data tipe semantik dan rule
2	Katakerja	Menyimpan data kata kerja dan tipe semantiknya
3	Rule	Menyimpan data rule grammar
4	Generator label	Menyimpan data generator label

Pada Tabel 2 dapat dilihat nama-nama tabel yang digunakan dalam sistem pelabelan peran semantis yang dibangun dalam penelitian ini. Tabel Semantik menyimpan data tipe semantik dan *rule*. Tabel Katakerja menyimpan data kata kerja dan tipe semantiknya. Tabel Rule menyimpan data *rule grammar*. Sedangkan tabel Generatorlabel menyimpan data generator label.

Secara umum, proses pelabelan peran semantis berdasarkan tata bahasa kasus dilakukan dengan langkah-langkah sebagai berikut:

1. Identifikasi kerangka kasus
2. Melabeli setiap kalimat dengan elemen kerangka yang sudah teridentifikasi

Ilustrasi implementasi pelabelan peran semantis yang diimplementasikan pada penelitian untuk bahasa Indonesia ini didasarkan pada teori tata bahasa kasus. Sebagai contoh, jika terdapat kalimat “Dia menulis surat lamaran” adalah sebagai berikut:

1. Setelah melalui proses penguraian sintaksis, diperoleh pohon urai sebagai berikut:

```

KALIMAT_1
|
KLAKTIF_2
|
SUBJ_3 PRED_6 OBJ_9
| | |
FPRON_4 FVPRED_7 FN_10
| | |
Pron_5 Vtran_8 NMINORG_11
WBLKGFN_13
dia menulis | |
N_12 NATR_14
surat |
NMINORG_15
|
N_16
lamaran

```

2. Dari pohon urai tersebut diperoleh peran sintaksis unsur-unsur kalimatnya, yaitu: “Dia” sebagai subjek, “menulis” sebagai predikat, dan “surat lamaran” sebagai objek.
3. Setelah diketahui peran sintaksis dari unsur-unsur kalimatnya, kemudian dilakukan pengecekan terhadap predikatnya untuk memperoleh tipe kata kerja, tata bahasa kasus dan *rule* yang sesuai, berdasarkan daftar kata kerja dan tipe semantiknya. Dari sini diperoleh informasi bahwa untuk predikat “menulis”, *rule*-nya adalah 8 dengan kerangka kasusnya adalah “+ [- A, O]”. Penentuan *rule* dan kerangka kasus dilakukan berdasarkan Tabel 2.1
4. Selanjutnya dilakukan pengisian *slot*, sehingga diperoleh hasil sebagai berikut: “Dia” sebagai A “menulis” sebagai KK “surat lamaran” sebagai O

5. Langkah selanjutnya adalah menentukan label peran sesuai dengan *rule* dan kerangka kasus yang telah ditemukan, berdasarkan pada aturan berikut:
 - a. Jika subjek merupakan AGENT (A), maka subjek diberi label ARG0, kemudian:
 - komponen setelah KK diberi label ARG1 dan ARG2 secaraurut untuk objek atau pelengkap yang ada pada kalimat
 - komponen setelah KK diberi label ARG-M untuk komponen/unsur yang menyatakan keterangan
 - b. Jika subjek bukan merupakan AGENT (A), maka subjek diberi label ARG1, kemudian
 - komponen setelah KK diberi label ARG2 untuk objek atau pelengkap yang ada pada kalimat, secaraurut
 - komponen setelah KK diberi label ARG-M untuk komponen atau unsur yang menyatakan keterangan

Dari langkah 5 diperoleh hasil sebagai berikut:

[ARG0 dia] [TARGET menulis] [ARG1 surat lamaran].

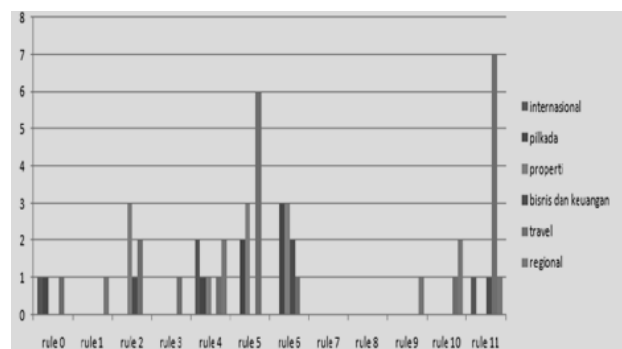
Kalimat yang telah diberi label peran semantis (menunjukkan struktur *verb-argument*) tersebut kemudian disimpan dalam *file* berekstensi .txt, dan kemudian digunakan sebagai masukan dari sistem pemodelan berbasis konsep.

Hasil Pengujian Sistem Pelabelan Peran Semantis

Pengujian dalam penelitian ini dilakukan untuk mengetahui keakuratan hasil pelabelan peran semantis yang dilakukan oleh sistem yang telah dibangun. Pengujian dibagi menjadi dua tahap. Pengujian tahap pertama dilakukan untuk mengetahui kemampuan sistem dalam mengklasifikasikan kata kerja

dalam setiap kalimat sesuai dengan tipe semantiknya. Pengujian tahap kedua dilakukan untuk mengetahui kemampuan sistem dalam melakukan pelabelan peran semantis secara terkomputerisasi, sesuai dengan teori tata bahasa kasus, di mana label peran semantis yang digunakan disesuaikan dengan basis data PropBank. Adapun analisis yang dilakukan adalah analisis terhadap hasil pelabelan peran semantis untuk setiap kata dalam kalimat dan membandingkan hasilnya dengan pelabelan secara manual berdasarkan basis data PropBank untuk bahasa Inggris.

Pada pengujian tahap pertama, kalimat masukkan yang diuraikan dengan menggunakan alat bantu PC-PATR diberi label sesuai dengan peran semantisnya. Kalimat masukkan yang berhasil diuraikan dengan menggunakan PC-PATR, kemudian diberi label sesuai dengan *peran semantisnya*. *Statistik hasil pelabelan peran semantis* dapat dilihat pada Gambar 4. Sumbu vertikal menunjukkan *rule* yang menggambarkan tipe kata kerja dalam *case grammar*. Sedangkan sumbu horisontal menunjukkan jumlah kalimat untuk masing-masing *rule* yang dapat dilihat menurut topiknya.



Gambar 4. Hasil Pengujian Tahap 2

Dari grafik hasil pengujian tahap 2 diperoleh informasi sebagai berikut:

1. Kalimat dengan tipe semantik KK Keadaan (*rule 0*) sebanyak 3 (tiga) buah, di mana 1 di antaranya berasal dari artikel berita dengan topik internasional,

- 1 (satu) buah kalimat berasal dari artikel berita dengan topik pilkada, dan 1 (satu) buah kalimat berasal dari artikel berita dengan topik travel.
2. Kalimat dengan tipe semantik KK Keadaan Pengalaman (*rule 1*) sebanyak 1 (satu) buah, yang berasal dari artikel berita dengan topik regional.
 3. Kalimat dengan tipe semantik KK Keadaan Benefaktif (*rule 2*) sebanyak 6 (enam) buah, di mana 3 (tiga) buah kalimat diantaranya berasal dari artikel berita dengan topik properti, 1 (satu) buah kalimat berasal dari artikel berita dengan topik bisnis dan keuangan, serta 2 (dua) buah kalimat lainnya berasal dari artikel berita dengan topik travel.
 4. Kalimat dengan tipe semantik KK Keadaan Lokatif (*rule 3*) sebanyak 1 (satu) buah, yang berasal dari artikel berita dengan topik travel.
 5. Kalimat dengan tipe semantik KK Proses (*rule 4*) sebanyak 4 (empat) buah, di mana 2 (dua) buah kalimat diantaranya berasal dari artikel berita dengan topik internasional, 2 (dua) buah kalimat berasal dari artikel berita dengan topik pilkada, dan 1 (satu) kalimat lainnya berasal dari artikel berita dengan topik properti.
 6. Kalimat dengan tipe semantik KK Aksi Pengalaman (*rule 9*) sebanyak 1 (satu) buah, yang berasal dari artikel berita dengan topik regional.
 7. Kalimat dengan tipe semantik KK Aksi Benefaktif (*rule 10*) sebanyak 3 (tiga) buah, yaitu 2 (dua) buah kalimat berasal dari artikel berita dengan topik regional dan sisanya berasal dari topik internasional.
 8. Kalimat dengan tipe semantik KK Aksi Lokatif (*rule 11*) sebanyak 9 (sembilan) buah kalimat, 7 (tujuh) diantaranya berasal dari artikel berita dengan topik travel, 1 (satu) buah kalimat berasal dari

artikel berita dengan topik bisnis dan keuangan, serta 1 (satu) buah kalimat lagi berasal dari artikel berita dengan topik regional.

9. Kalimat dengan tipe semantik KK Proses Pengalaman (*rule 5*) sebanyak 11 (sebelas) buah, di mana 6 (enam) diantaranya berasal dari artikel berita dengan topik travel, 3 (tiga) buah kalimat berasal dari artikel berita dengan topik properti, dan sisanya berasal dari artikel berita dengan topik pilkada.
10. Kalimat dengan tipe semantik KK Proses Benefaktif (*rule 6*) sebanyak 9 (sembilan) buah, di mana 3 (tiga) diantaranya berasal dari artikel berita dengan topik pilkada, 3 (tiga) buah berasal dari artikel berita dengan topik properti, 2 (dua) buah kalimat berasal dari artikel berita dengan topik bisnis dan keuangan, dan sisanya berasal dari artikel berita dengan topik travel.

Jumlah keseluruhan kalimat yang digunakan dalam pengujian pada tahap kedua adalah 52 (lima puluh dua) kalimat. Dari 52 (lima puluh dua) kalimat tersebut, 4 (empat) di antaranya berasal dari artikel berita dengan topik internasional, 7 (tujuh) kalimat berasal dari artikel berita dengan topik pilkada, 10 (sepuluh) kalimat berasal dari artikel berita dengan topik properti, 4 (empat) kalimat berasal dari artikel berita dengan topik bisnis dan keuangan, 20 (dua puluh) kalimat berasal dari artikel berita dengan topik travel, dan sisanya 7 (tujuh) kalimat, berasal dari artikel berita dengan topik regional.

Hasil pelabelan peran semantis yang dilakukan berdasarkan teori tata bahasa kasus dapat mengklasifikasikan semua kata kerja berdasarkan tipe semantiknya dan dapat memberikan label peran semantis dengan benar pada setiap kata yang terdapat pada kalimat-kalimat tersebut sesuai dengan ketentuan pelabelan peran semantis pada basis data PropBank. Dari 48 (empat puluh delapan) kalimat yang digunakan pada pengujian tahap

pertama dan 52 (lima puluh dua) kalimat yang digunakan pada pengujian tahap kedua, semua kata kerja dapat diklasifikasikan dengan benar sesuai dengan tipe semantiknya dan dapat diberikan label peran semantis sesuai dengan basis data PropBank (Palmer, Kingsbury & Gildea, 2002).

Analisis Hasil Pelabelan Peran Semantis

Setelah dibandingkan dengan hasil pelabelan secara manual, tidak terdapat kesalahan dalam pelabelan peran semantis untuk setiap kata dalam kalimat. Untuk menentukan benar atau tidaknya hasil pelabelan, dilakukan dengan membandingkan kata yang terlabeli dengan benar sesuai peran semantisnya. Sebagai contoh, jika diketahui hasil pelabelan dari sistem:

[ARG0 para murid] [TARGET mengajukan] [ARG1 pertanyaan] [ARG-2 kepada guru]

Sedangkan hasil pelabelan secara manual adalah:

[ARG0 para murid] [TARGET mengajukan] [ARG1 pertanyaan] [ARG-M kepada guru]

maka kata “kepada guru” disebut sebagai hasil pelabelan yang salah dari sistem.

Analisis pelabelan untuk masing-masing tipe kata kerja dasar (pada penelitian ini digunakan 12 (dua belas) tipe kata kerja dasar berdasarkan teori tata bahasa kasus adalah sebagai berikut (contoh untuk masing-masing kata kerja pada pengujian tahap pertama):

1. Kata kerja keadaan

Kasus atau *frame element* yang diperlukan pada tipe kata kerja ini adalah Objek keadaan (Ok), sehingga *case frame*-nya adalah + [- Ok]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label

yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Perempuan itu cantik.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari tata bahasa kasus, yaitu sebagai berikut:

“Perempuan itu” sebagai Ok

“cantik” sebagai KK

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, maka ditentukan label peran semantis sesuai dengan basis data PropBank, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 perempuan itu] [TARGET cantik]

2. Kata kerja keadaan-pengalam

Kasus yang diperlukan pada tipe kata kerja ini adalah Pengalam (P) dan Objek keadaan (Ok), sehingga *case frame*-nya adalah + [- P, Ok]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Dia mampu hidup mandiri.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Dia” sebagai P

“mampu” sebagai KK

“hidup mandiri” sebagai Ok

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 dia] [TARGET mampu]
[ARG2 hidup mandiri]

3. Kata kerja keadaan-benefaktif

Kasus yang diperlukan pada tipe kata kerja ini adalah Benefaktif (B) dan Objek keadaan (Ok), sehingga *case frame*-nya adalah + [- B, Ok]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Pertunjukan teater itu memuaskan penonton.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Pertunjukan teater itu” sebagai B
“memuaskan” sebagai KK
“penonton” sebagai Ok

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 pertunjukan teater itu]
[TARGET memuaskan] [ARG2 penonton]

4. Kata kerja keadaan-lokatif

Kasus yang diperlukan pada tipe kata kerja ini adalah Objek keadaan (Ok) dan Lokatif (L), sehingga *case frame*-nya adalah + [- Ok, L]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Mobil itu berhenti di tepi jalan.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Mobil itu” sebagai Ok
“berhenti” sebagai KK
“di tepi jalan” sebagai L

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 mobil itu] [TARGET berhenti]
[ARG-M di tepi jalan]

5. Kata kerja proses

Kasus yang diperlukan pada tipe kata kerja ini adalah Objek (O), sehingga *case frame*-nya adalah + [- O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Asap Gunung Merapi itu terus mengepul.
Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Asap Gunung Merapi itu” sebagai O
“terus mengepul” sebagai KK

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 asap gunung merapi itu]
[TARGET terus mengepul]

6. Kata kerja proses-pengalam

Kasus yang diperlukan pada tipe kata kerja ini adalah Pengalam (P) dan Objek (O), sehingga *case frame*-nya adalah + [- P, O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Nelayan itu mendengar suara pesawat

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Nelayan itu” sebagai P
 “mendengar” sebagai KK
 “suara pesawat” sebagai O

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 nelayan itu] [TARGET mendengar] [ARG2 suara pesawat]

7. Kata kerja proses-benefaktif

Kasus yang diperlukan pada tipe kata kerja ini adalah Benefaktif (B) dan Objek (O), sehingga *case frame*-nya adalah + [- B, O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Pejabat itu beristri penyanyi dangdut.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Pejabat itu” sebagai B
 “beristri” sebagai KK
 “penyanyi dangdut” O

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 pejabat itu] [TARGET beristri]
 [ARG2 penyanyi dangdut]

8. Kata kerja proses-lokatif

Kasus yang diperlukan pada tipe kata kerja ini adalah Objek (O) dan Lokatif (L), sehingga *case frame*-nya adalah + [- O, L]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Anak kecil itu hanyut di sungai.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Anak kecil itu” sebagai O
 “hanyut” sebagai KK
 “di sungai” sebagai L

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 anak kecil itu] [TARGET hanyut]
 [ARG-M di sungai]

9. Kata kerja aksi

Kasus yang diperlukan pada tipe kata kerja ini adalah Agen (A) dan Objek (O), sehingga *case frame*-nya adalah + [- A, O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

[ARG0 dia] [TARGET menulis]
[ARG1 surat lamaran]

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Dia” sebagai A
“menulis” sebagai KK
“surat lamaran” sebagai O

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG0 dia] [TARGET menulis]
[ARG1 surat lamaran]

10. Kata kerja aksi-pengalam

Kasus yang diperlukan pada tipe kata kerja ini adalah Agen (A), Pengalaman (P), dan Objek (O), sehingga *case frame*-nya adalah + [- A, P, O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Pak guru menjelaskan materi pelajaran kepada murid.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Pak guru” sebagai A
“menjelaskan” sebagai KK
“materi pelajaran” sebagai P
“kepada murid” sebagai O

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG0 pak guru] [TARGET menjelaskan]
[ARG1 materi pelajaran]
[ARGM kepada murid]

11. Kata kerja aksi-benefaktif

Kasus yang diperlukan pada tipe kata kerja ini adalah Agen (A), Benefaktif (B) dan Objek (O), sehingga *case frame*-nya adalah + [- A, B, O]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

[ARG0 mahasiswa itu] [TARGET mendapatkan]
[ARG1 beasiswa]

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Mahasiswa itu” sebagai A
“mendapatkan” sebagai KK
“beasiswa” sebagai O

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG0 mahasiswa itu] [TARGET mendapatkan]
[ARG1 beasiswa]

12. Kata kerja aksi-lokatif

Kasus yang diperlukan pada tipe kata kerja ini adalah Agen (A), Objek (O), dan Lokatif (L), sehingga *case frame*-nya adalah + [- A, O, L]. Jika telah ditemukan *case frame*-nya, maka akan di-generate label peran semantis sesuai dengan label yang ada pada basis data PropBank. Misalnya untuk kalimat berikut:

Adik mengembalikan buku ke perpustakaan.

Setelah diperoleh *case frame*-nya, maka akan dilakukan pengisian *slot* dari *case grammar*, yaitu sebagai berikut:

“Adik” sebagai A
 “mengembalikan” sebagai KK
 “buku” sebagai O
 “ke perpustakaan” sebagai L

Kemudian berdasarkan informasi mengenai *case frame* dan struktur sintaksisnya, dengan tetap memperhatikan aspek sintaksis, sehingga diperoleh hasil sebagai berikut:

[ARG1 anak kecil itu] [TARGET hanyut]
 [ARG-M di sungai]

PENUTUP

Pelabelan peran semantis kalimat berbahasa Indonesia, dengan menggunakan tata bahasa kasus dapat mengatasi permasalahan sumber daya manusia dan waktu yang diperlukan untuk membangun basis data *PropBank*. Akan tetapi, cakupan dari kalimat yang dapat terlabeli dengan benar menggunakan tata bahasa kasus masih terbatas pada kalimat deklaratif dan kalimat tunggal, di mana pola kalimatnya juga terbatas pada pola-pola tertentu saja. Keterbatasan kalimat yang dapat dilabeli disebabkan belum tersedianya pengurai kalimat berbahasa Indonesia yang memadai untuk menguraikan kalimat majemuk untuk semua pola, kalimat tanya, dan kalimat seru.

Implementasi pelabelan peran semantis kalimat bahasa Indonesia dengan menggunakan tata bahasa kasus dapat menyelesaikan masalah pelabelan peran semantis untuk kalimat tunggal dan kalimat deklaratif untuk pola-pola tertentu, secara terkomputerisasi. Hasil implementasi pelabelan peran semantis kalimat bahasa Indonesia dengan menggunakan tata bahasa

kasus memberikan hasil yang akurat untuk pelabelan kalimat tunggal bahasa Indonesia, di mana kata kerjanya telah diketahui tipe semantiknya, jenis kalimatnya adalah kalimat deklaratif dan pola kalimatnya tertentu saja.

Pelabelan peran semantis timbul dari kebutuhan akan pemahaman dan pemanipulasian teks untuk mendapatkan informasi tertentu dari sekumpulan dokumen teks dengan memanfaatkan teknologi komputer namun tetap memperhatikan aspek sintaksis dan semantis, mengingat semakin besarnya volume berita elektronik berbahasa Indonesia yang di dalamnya terdapat berbagai informasi bermanfaat namun tidak bisa diperoleh secara langsung. Seiring dengan semakin besarnya volume berita dan semakin diperlukannya teknik untuk mendapatkan informasi berharga dari suatu artikel berita, perlu dilakukan penelitian untuk membantu pelabelan peran semantis dengan memanfaatkan teknologi komputer, yang pada akhirnya dapat membantu pembaca artikel berita atau dokumen teks lainnya untuk memperoleh informasi-informasi berharga.

Pelabelan peran semantis kalimat tunggal berbahasa Indonesia memberikan dasar bagi pengembangan pelabelan peran semantis untuk kalimat majemuk berbahasa Indonesia. Mengidentifikasi peran semantis dapat memberikan level analisis semantis yang bermanfaat dalam menyelesaikan permasalahan pengolahan bahasa manusia. Oleh karena itu, diperlukan penelitian lebih lanjut mengenai pengurai sintaksis untuk kalimat majemuk dan pelabelan peran semantis kalimat berbahasa Indonesia, sehingga tidak hanya terbatas pada kalimat tunggal dan kalimat deklaratif saja, tetapi juga kalimat majemuk dan jenis kalimat lainnya seperti kalimat seru dan kalimat tanya. Selain itu, penelitian lebih lanjut perlu dilakukan agar dapat dihasilkan sistem pelabelan peran semantis untuk kalimat berbahasa Indonesia yang terintegrasi dengan sistem penandaan *part-of-speech* dan pengurai sintaksis.

Pada penelitian lanjutan, dapat dikembangkan dan diterapkan konsep pelabelan

peran semantis dalam implementasi *e-office* maupun *e-government*. Konsep pelabelan peran semantis dapat dikembangkan dan diterapkan dalam pada aplikasi *e-office* atau *e-government*. Sebagai contoh, dengan menerapkan konsep pelabelan peran semantis, dapat dilakukan pengelompokan surat elektronik, nota dinas, atau dokumen lainnya yang diterima atau dikirim, misalnya apakah dikategorikan *spam* atau *non-spam*, dengan memperhatikan unsur semantis sehingga diperoleh hasil yang lebih akurat dibandingkan pengelompokan dokumen atau surat elektronik tanpa memperhatikan unsur semantis. Pada *e-office* atau *e-government* yang sudah menerapkan fitur pengelolaan dokumen, konsep pelabelan peran semantis dapat dikembangkan dan diterapkan dalam meningkatkan keakuratan pengelompokan dan pencarian surat, nota dinas, atau dokumen elektronik lainnya.

DAFTAR PUSTAKA

- Albright, E. (2006). *An Overview of historical development of semantic role representation*.
- Gildea, D., Jurafsky, D. (2002). *Automatic Labeling of Semantic Roles*. (<http://www.cs.rochester.edu/~gildea/gildea-cl02.pdf>, diperoleh tanggal 16 Oktober 2008).
- Gildea, D., Palmer, M. (2002). *The Necessity of parsing for predicate argument Recognition*. (<http://acl.ldc.upenn.edu/P/P02/P02-1031.pdf>, diperoleh tanggal 15 Oktober 2008).
- Musfirah, U. (2007). *Interpretasi semantik kalimat bahasa Indonesia dalam bentuk tanya jawab menggunakan Case Grammar*. Bandung: Institut Teknologi Telkom.
- Palmer, M. , Kingsbury, P., & Gildea, D. (2002). *The Proposition Bank: An Annotated corpus of semantic roles*. (<http://www.cs.rochester.edu/~gildea/palmer-propbank-cl.pdf>, diperoleh tanggal 2 November 2008).
- Sawitri, A. D. (2007). Penentuan peran semantis kata dalam kalimat bahasa Indonesia menggunakan *Case Grammar*. Bandung: Institut Teknologi Telkom.
- Shehata, S., Karray, F., & Kamel, M. (2005). *Concept mining using Conceptual Ontological Graph (COG)*.
- Shehata, S., Karray, F., & Kamel, M. (2006). Enhancing Text Retrieval Performance using Conceptual Ontological Graph. *Proceedings of ICDMW 2006*, 39-44.
- Shehata, S., Karray, & F., Kamel, M. (2007). A Concept-based Model for Enhancing Text Categorization. *Proceedings of ICDMW 2007*, 12-15.
- Tarigan, H. G. (1990). *Pengajaran tata bahasa kasus*. Bandung: Penerbit Angkasa.
- Wibisono, Y., & Khodra, M. L. (2006). *Clustering berita berbahasa Indonesia*.